A New Rating Scale for Age-Related White Matter Changes Applicable to MRI and CT

L.O. Wahlund, MD, PhD; F. Barkhof, MD, PhD; F. Fazekas, MD; L. Bronge, MD; M. Augustin, MD; M. Sjögren, MD, PhD; A. Wallin, MD, PhD; H. Ader, PhD; D. Leys, MD, PhD; L. Pantoni, MD, PhD; M. Pasquier, MD, PhD; T. Erkinjuntti, MD, PhD; P. Scheltens, MD, PhD; on behalf of the European Task Force on Age-Related White Matter Changes

Background and Purpose—MRI is more sensitive than CT for detection of age-related white matter changes (ARWMC). Most rating scales estimate the degree and distribution of ARWMC either on CT or on MRI, and they differ in many aspects. This makes it difficult to compare CT and MRI studies. To be able to study the evolution and possible effect of drug treatment on ARWMC in large patient samples, it is necessary to have a rating scale constructed for both MRI and CT. We have developed and evaluated a new scale and studied ARWMC in a large number of patients examined with both MRI and CT.

Methods—Seventy-seven patients with ARWMC on either CT or MRI were recruited and a complementary examination (MRI or CT) performed. The patients came from 4 centers in Europe, and the scans were rated by 4 raters on 1 occasion with the new ARWMC rating scale. The interrater reliability was evaluated by using $\kappa$ statistics. The degree and distribution of ARWMC in CT and MRI scans were compared in different brain areas.

Results—Interrater reliability was good for MRI ($\kappa=0.67$) and moderate for CT ($\kappa=0.48$). MRI was superior in detection of small ARWMC, whereas larger lesions were detected equally well with both CT and MRI. In the parieto-occipital and infratentorial areas, MRI detected significantly more ARWMC than did CT. In the frontal area and basal ganglia, no differences between modalities were found. When a fluid-attenuated inversion recovery sequence was used, MRI detected significantly more lesions than CT in frontal and parieto-occipital areas. No differences were found in basal ganglia and infratentorial areas.

Conclusions—We present a new ARWMC scale applicable to both CT and MRI that has almost equal sensitivity, except for certain regions. The interrater reliability was slightly better for MRI, as was the detectability of small lesions. (Stroke. 2001;32:1318-1322.)

Key Words: dementia ■ magnetic resonance imaging ■ rating scales ■ tomography, x-ray computed ■ white matter

White matter changes (WMC) are defined as areas with high signal intensities on T2-weighted MRI and as areas with low attenuation on CT. The mechanisms for development of WMC are not fully understood, but several histopathologic correlates have been reported. These include enlarged perivascular (Virchow-Robin) spaces, as well as degeneration of myelin and axons with increased intracellular and extracellular water content, gliosis, and even infarction.1–10

The clinical significance of WMC has not been fully elucidated. There is a relationship between several cerebrovascular risk factors and the presence of WMC. One of the strongest risk factors, apart from hypertension, is that of age.11–13 Henceforth, we will designate WMC as “age-related white matter changes” (ARWMC).

There is also evidence for a relationship between ARWMC and cognitive impairment in demented patients14–16 as well as in healthy elderly individuals.13,17–19 However, the extent of this association is still controversial. Little is known about the natural history of ARWMC, starting from their development to possibly subsequent progression. It is also unclear whether it is possible to affect the evolution of ARWMC with...
pharmacological treatment and whether this would have any impact on cognitive performance or other tasks that require more complex cerebral processing, such as coordinated movement. The advent of MRI has focused attention on ARWMC due to the conspicuousness of such lesions on proton density (PD) and T2-weighted images, including fluid-attenuated inversion recovery (FLAIR). Also, the high sensitivity of this technique has increasingly brought into question the utility of CT for delineating ARWMC.

Given the possible effect of ARWMC on cognitive function in modern pharmacological antidementia treatment, there is a need for the evaluation of ARWMC. The degree and distribution of ARWMC are important in vascular dementia as well as in other dementia disorders, such as Alzheimer’s disease. CT is less costly and more easily performed than MRI, and in many regions the number of MRI scanners is still limited. Several rating scales exist for visual rating of ARWMC on MRI scans. Many of them are validated and widely used. Similarly, there are a number of rating scales for CT. At present, there is only a single study in which the authors have attempted to design a scale applicable to both CT and MRI scans. However, this scale, as designed by van Swieten, uses different criteria for CT and MRI scans, thus leading to different sensitivities for the respective modalities. It is therefore impossible to compare the results of different studies on ARWMC within and, even more so, between the 2 imaging modalities. By generating a combined CT-MRI ARWMC rating scale, we wanted to provide a tool for such between-modality comparison as well as to have an instrument for comparing the sensitivity for ARWMC and the reliability of their rating between CT and MRI.

Such a scale was developed under the auspices of the European Task Force on Age-Related White Matter Changes. In this study, we compared the regional sensitivity of MRI and CT for ARWMC within the same patient. The interobserver reliability of the new scale in both imaging techniques was also studied.

### Subjects and Methods

In the ARWMC scale, the degree of white matter changes is rated on a 4-point scale. Ratings were done on CT images by using standard hard copies and on MRI images on computer screen with either PD and T2-weighted images or T2 and FLAIR images. Examples of the different rating scores are shown in Figure 1.

White matter changes on MRI were defined as ill-defined hyperintensities ≥5 mm on both T2 and PD/FLAIR images, and on CT as ill-defined and moderately hypodense areas of ≥5 mm. Lacunes were defined as well-defined areas of ≥2 mm with attenuation (on CT) or signal characteristics (on MRI) the same as cerebrospinal fluid. If lesions with these characteristics were ≥2 mm, they were considered perivascular spaces, except around the anterior commissure, where perivascular spaces can be large.

Changes in the basal ganglia were rated in the same way and considered white matter lesions even if located in the gray matter nuclei, which contains a small amount of white matter. The definitions of rating scores (0–3) are shown in Table 1.

Five different regions were rated in the right and left hemispheres separately: (1) the frontal area, which was the frontal lobe anterior to the central sulcus; (2) the parieto-occipital area, which consisted of the parietal and occipital lobes together; (3) the temporal area, which was the temporal lobe (the border between the parieto-occipital and temporal lobes was approximated as a line drawn from the posterior part of the Sylvian fissure to the trigone areas of the lateral ventricles); (4) the infratentorial area, which included the brain stem and cerebellum; and (5) the basal ganglia, which included the striatum, globus pallidus, thalamus, internal and external capsules, and insula.

All ratings were performed on one occasion at Huddinge Hospital. After a training session in which all 4 raters (F.B., L.B., M.A., and M.S.) together rated 10 cases, the raters were divided into pairs. Each pair of raters then evaluated MRI and CT scans from half of the cases. Within each pair, the 2 raters reached consensus. Inter-rater reliability values were calculated from a further series of rating both MRI and CT images from 20 patients, in which all 4 raters rated individually.

Previously acquired cerebral scans from 77 patients were used in this study. The patients had been examined at either Huddinge (n = 20), Amsterdam (n = 5), Lille (n = 14), or Graz (n = 38). They were included if ARWMC was noticed on a routine examination (on either MRI or CT) and when both MRI and CT had been performed no more than 3 months apart. We assume that no major alterations in the degree of white matter changes occurred during this time period. Patients with other gross pathological findings such as tumors, large bleedings, and recent territorial infarctions were excluded. We did not consider the final diagnosis or underlying risk factors in these patients; our only purpose was to compare ARWMC on MRI and CT scans. To simulate routine clinical conditions, we also did not advise...
TABLE 1. The ARWMC Rating Scale for MRI and CT

<table>
<thead>
<tr>
<th>White matter lesions</th>
<th>MRI</th>
<th>CT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No lesions (including symmetrical, well-defined caps or bands)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Focal lesions</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Beginning confluence of lesions</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Diffuse involvement of the entire region, with or without involvement of U fibers</td>
<td></td>
</tr>
</tbody>
</table>

Basal ganglia lesions

<table>
<thead>
<tr>
<th>Location</th>
<th>MRI</th>
<th>CT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frontal</td>
<td>0.97±0.99</td>
<td>0.95±1.12</td>
</tr>
<tr>
<td>Parieto-occipital</td>
<td>1.20±1.11</td>
<td>1.01±1.21</td>
</tr>
<tr>
<td>Temporal</td>
<td>1.17±1.15</td>
<td>1.04±1.12</td>
</tr>
<tr>
<td>Basal ganglia</td>
<td>0.23±0.56</td>
<td>0.31±0.63</td>
</tr>
<tr>
<td>Infratentorial</td>
<td>0.24±0.63</td>
<td>0.26±0.69</td>
</tr>
</tbody>
</table>

Statistics

No descriptive data were available to perform power calculations. Therefore, type II errors could not be excluded for some comparisons.

The nonparametric sign test was used to compare visual ratings between MRI and CT. Weighted \( \kappa \) values were calculated as a measure of reproducibility: \( \kappa \) values <0.4 indicated poor agreement, values from 0.41 to 0.6 indicated moderate agreement, values from 0.61 to 0.8 indicated good agreement, and \( \kappa \) values >0.81 indicated excellent agreement.21

Results

The means and standard deviations of the MRI and CT ratings in each separate area are shown in Table 2 (medians are not shown due to lack of variability). As expected, most lesions were found in the frontal and parieto-occipital regions. The distribution of the rating scores (sum of the right and left side) according to brain region is presented in Figure 2. There was a size-dependent difference in the distribution of rating scores between MRI and CT. MRI detected more of the small ARWMC whereas CT was equal or superior in detecting larger lesions.

To address the relative detection capacity of each imaging modality, we compared in a pairwise fashion the MRI and CT and described them as either (1) MRI and CT scoring equal, (2) MRI scoring higher than CT, or (3) CT scoring higher than MRI. Few lesions were detected in the temporal areas, and therefore no comparison was made between MRI and CT in this region. The data are shown as a percentage of all rated subjects. As evident from Table 3, MRI and CT were equal in >50% of all ratings. However, in the parieto-occipital and infratentorial areas, MRI rated significantly more ARWMC than did CT. When the MRI ratings were based on the FLAIR sequence (38 subjects), significantly more ARWMC were rated in all areas except for the basal ganglia and infratentorial area. Also in these cases, >50% of all ratings from CT were equal to MRI (see Table 4).

The results of the interrater reliability are presented in Table 4. This showed moderate-excellent agreement for MRI and CT. The highest \( \kappa \) values were found for the rating of frontal, parieto-occipital, and basal ganglia areas, whereas the \( \kappa \) score was lower for the temporal area as well as for the number of infarcts. As expected, the infratentorial area was significantly more reliably rated with MRI than with CT.

When ARWMC in all areas were considered together, good agreement was found for MRI (0.67) and moderate agreement for CT (0.48). From Table 5 it is evident that the \( \kappa \) value was larger for MRI than CT in all but 1 region. The lack of statistical significance might be due to type II error.

Discussion

To be able to assess the degree of ARWMC from both MRI and CT examinations obtained on a clinical, day-to-day basis, we developed a scale that can be applied to MRI and CT alike. This ARWMC scale could also be used in clinical trials with large patient samples, in which drugs affecting brain degenerative processes are tested and both MRI and CT are used for evaluation.

The scale was evaluated by 4 raters and showed good reliability for both CT and MRI in most regions. As expected, the reliability was higher for MRI rating than for CT rating. This was especially true in the frontal and parieto-occipital regions, where very high \( \kappa \) values were detected. Furthermore, MRI was significantly more reliably rated in the parietal/occipital area than CT. This might be explained by...
the fact that CT scans were obtained from a variety of scanners with different slice thickness and image quality.

We found that when using standard T2/PD images, MRI was more sensitive in detecting ARWMC in the parieto-occipital and infratentorial regions of the brain. The latter region is easier to study with MRI than with CT because no bone artifacts are present. Moreover, the differences were mainly due to the ability of MRI to detect small ARWMCs. CT was as good as MRI in detecting larger lesions. This might explain why many studies have shown that ARWMC found on CT correlate better with symptoms than do the changes detected with MRI. When the FLAIR sequence was used for MRI rating, ARWMC were found significantly more often than with CT. This suggests that FLAIR may be preferable for the study of ARWMC. However, this issue will have to be addressed more extensively in a further study.

To summarize, this study shows that differences of MRI and CT in detecting ARWMC are primarily related to lesion size. MRI was superior due to its better detection of small ARWMC, whereas medium and large lesions were detected

<table>
<thead>
<tr>
<th>TABLE 3. Comparison of MRI and CT in Detection of ARWMC, by Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area</td>
</tr>
<tr>
<td>-----------------------</td>
</tr>
<tr>
<td>Frontal</td>
</tr>
<tr>
<td>Parieto-occipital</td>
</tr>
<tr>
<td>Basal ganglia</td>
</tr>
<tr>
<td>Infratentorial</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 4. Comparison of MRI and CT in Detection of ARWMC in the Subsample of 38 Patients With FLAIR Images</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area</td>
</tr>
<tr>
<td>-----------------------</td>
</tr>
<tr>
<td>Frontal</td>
</tr>
<tr>
<td>Parieto-occipital</td>
</tr>
<tr>
<td>Basal ganglia</td>
</tr>
<tr>
<td>Infratentorial</td>
</tr>
</tbody>
</table>

Based on ratings of all 77 subjects. The columns show percentage of patients in whom MRI scored higher than CT, CT scored higher than MRI, and CT and MRI scored equal. P values obtained sign test.

TABLE 4. Comparison of MRI and CT in Detection of ARWMC in the Subsample of 38 Patients With FLAIR Images

The columns show percentage of total number of ratings in which MRI scored higher than CT, CT scored higher than MRI, and CT and MRI scored equal. P values obtained with sign test.
TABLE 5. \( \kappa \) Scores for Interobserver Agreement or Rating of ARWMC

<table>
<thead>
<tr>
<th>Location</th>
<th>MRI</th>
<th>CT</th>
<th>( P^* )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frontal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>0.8512 (0.0496)</td>
<td>0.7238 (0.0709)</td>
<td>0.1409</td>
</tr>
<tr>
<td>L</td>
<td>0.8622 (0.0447)</td>
<td>0.7215 (0.0718)</td>
<td>0.0962</td>
</tr>
<tr>
<td>Total</td>
<td>0.8560 (0.0401)</td>
<td>0.7723 (0.0693)</td>
<td>0.2958</td>
</tr>
<tr>
<td>Parieto-occipital</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>0.8680 (0.0386)</td>
<td>0.7281 (0.0629)</td>
<td>0.058</td>
</tr>
<tr>
<td>L</td>
<td>0.8831 (0.0399)</td>
<td>0.6626 (0.0711)</td>
<td>0.0068</td>
</tr>
<tr>
<td>Total</td>
<td>0.8749 (0.0345)</td>
<td>0.7155 (0.0639)</td>
<td>0.0282</td>
</tr>
<tr>
<td>Temporal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>0.5841 (0.1008)</td>
<td>0.3474 (0.1188)</td>
<td>0.2308</td>
</tr>
<tr>
<td>L</td>
<td>0.6190 (0.0968)</td>
<td>0.6258 (0.0868)</td>
<td>0.9583</td>
</tr>
<tr>
<td>Total</td>
<td>0.6033 (0.0903)</td>
<td>0.5103 (0.0948)</td>
<td>0.3057</td>
</tr>
<tr>
<td>Basal ganglia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>0.6627 (0.0723)</td>
<td>0.4561 (0.0835)</td>
<td>0.0614</td>
</tr>
<tr>
<td>L</td>
<td>0.6613 (0.0773)</td>
<td>0.4827 (0.0678)</td>
<td>0.0824</td>
</tr>
<tr>
<td>Total</td>
<td>0.6535 (0.0698)</td>
<td>0.4707 (0.0667)</td>
<td>0.0583</td>
</tr>
<tr>
<td>Infratentorial</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>0.5858 (0.0944)</td>
<td>0.3093 (0.0167)</td>
<td>0.0039</td>
</tr>
<tr>
<td>L</td>
<td>0.5981 (0.0804)</td>
<td>0.0983 (0.0772)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Total</td>
<td>0.5826 (0.0818)</td>
<td>0.1845 (0.0633)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Frontal + temporal + parieto-occipital</td>
<td>0.6654 (0.089)</td>
<td>0.4837 (0.0907)</td>
<td>0.1527</td>
</tr>
</tbody>
</table>

Agreement is shown according to location for MRI and CT in 20 patients. Values in parentheses are SD. *P value of test for \( \kappa \) differences between MRI and CT.

equally well by both modalities. Regional differences played a subordinate role, which is not unexpected from the predominating location of ARWMC in areas that can usually be imaged with high quality by both MRI and CT. Therefore, CT studies in combination with MRI might be used in evaluation of ARWMC in multicenter studies when attempting to address primarily more marked white matter damage. (This could also be the type of white matter damage that shows closer relation to clinical-cognitive deficits, at least in cross-sectional studies.) In this context, the new ARWMC rating scale may be a useful tool that demonstrates good interrater reliability. However, for minor ARWMC, MRI appears mandatory, and it is probably also the technique needed for monitoring ARWMC progression.

Acknowledgments

This study was financially supported by Alzheimerfonden, Lund. We acknowledge Dr Richard Cowburn, research nurse Ulla Swahn, and Dr Leif Svensson for excellent typographic and technical support.

References


A New Rating Scale for Age-Related White Matter Changes Applicable to MRI and CT
on behalf of the European Task Force on Age-Related White Matter Changes

Stroke. 2001;32:1318-1322
doi: 10.1161/01.STR.32.6.1318

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/32/6/1318

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Stroke can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Stroke is online at:
http://stroke.ahajournals.org//subscriptions/